

# Projected shrinkage algorithm for box-constrained $\ell_1$ -minimization

Hui Zhang

LiZhi Cheng \*

August 25, 2014

## Abstract

Box-constrained  $\ell_1$ -minimization can perform remarkably better than classical  $\ell_1$ -minimization when correction box constraints are available. And also many practical  $\ell_1$ -minimization models indeed involve box constraints because they take certain values from some interval. In this paper, we propose an efficient iteration scheme, namely projected shrinkage (ProShrink) algorithm, to solve a class of box-constrained  $\ell_1$ -minimization problems. A key contribution in our technique is that a complicated proximal point operator appeared in the deduction can be equivalently simplified into a projected shrinkage operator. Theoretically, we prove that ProShrink enjoys a convergence of both the primal and dual point sequences. On the numerical level, we demonstrate the benefit of adding box constraints via sparse recovery experiments.

**Keywords:** proximal point operator; projected shrinkage; box constraints;  $\ell_1$ -minimization; sparse recovery

## 1 Introduction

The past two decades has witnessed the wide application of  $\ell_1$ -minimization models in signal and image processing, compressive sensing, machine learning, statistic, and more. The success of  $\ell_1$  minimization is mainly due to that the  $\ell_1$ -norm can well reflect sparse prior. Recently, it was observed that other auxiliary information of sparse solutions, such as partial support set [12] and nonnegative sparsity [1, 4], could help fit practical models. In this paper, instead of studying the theoretical benefit of modeling auxiliary priors, we are interested in designing efficient algorithms to solve the  $\ell_1$ -minimization problems with auxiliary box constraints:

$$\underset{x}{\text{minimize}} \|x\|_1, \quad \text{s.t. } Ax = b, x \in \mathcal{X} \quad (1)$$

and

$$\underset{x}{\text{minimize}} \|x\|_1 + \frac{1}{2\tau} \|x\|_2^2, \quad \text{s.t. } Ax = b, x \in \mathcal{X} \quad (2)$$

where  $A \in \mathbb{R}^{n \times m}$ ,  $b \in \mathbb{R}^m$  are given,  $\tau$  is an augmented parameter, and  $\mathcal{X}$  is some box-constrained set. The above problems are obtained separately by imposing box constraints to the basis pursuit model [3]:

$$\underset{x}{\text{minimize}} \|x\|_1, \quad \text{s.t. } Ax = b \quad (3)$$

and the augmented  $\ell_1$  norm model [5]:

$$\underset{x}{\text{minimize}} \|x\|_1 + \frac{1}{2\tau} \|x\|_2^2, \quad \text{s.t. } Ax = b, \quad (4)$$

---

\*College of Science, National University of Defense Technology, Changsha, Hunan, China, 410073. Emails: h.zhang1984@163.com( Hui Zhang); clzcheng@nudt.edu.cn(Lizhi Cheng).

both of which have been proved powerful for sparse recovery. Adding box constraints to classical  $\ell_1$  minimization on one hand extends the range of models (3) and (4) to include more practical models in application, and on the other hand can help improve the ability of sparse recovery of them when correct box constraints are available; the second point of view shall be demonstrated numerically later on. Similar benefit of adding box constraints to classical matrix completion has been observed in a recent paper [11] which was posted on arXiv at the time of the writing of the present paper.

Due to the existing of the strongly convex term  $\frac{1}{\tau}\|x\|_2^2$ , it has been explained in several papers [14, 15, 5] that models (2) and (4) have computational advantages over their correspondences (1) and (3). Besides, applying the proximal point algorithm [9] to models (1) or (3) generates a series of subproblems similar to (2) or (4). Therefore, the center assignment of solving problems (1)-(4) reduces to studying the following generalized problem

$$\underset{x}{\text{minimize}} \|x\|_1 + \frac{1}{2\tau}\|x - u\|_2^2, \quad \text{s.t. } Ax = b, x \in \mathcal{X} \quad (5)$$

where  $u$  is a given vector. With the help of the Lagrange dual analysis and by noticing the strong convexity of the objective function, in this study we derive a projected shrinkage (ProShink) algorithm for solving (5). By the Nesterov techniques [7], the proposed algorithm can be speeded up; we present an accelerated scheme as well. Theoretically, we prove the convergence of both the primal and dual point sequences of ProShink. A key contribution in our technique is that a complicated proximal operator appeared in the deduction can be equivalently simplified into a projected shrinkage operator. This can also be applied to simplifying standard forward-backward splitting algorithm for the boxed-constrained basis pursuit denoising problem:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \|x\|_1 + \frac{1}{2\lambda}\|Ax - b\|_2^2, \quad (6)$$

where  $\lambda$  is a positive paramter.

The rest of paper is organized as follows. In section 2, we introduce some basis concepts of constrained convex optimization and obtain important properties about the shrinkage operator. In section 3, under the Lagrange dual analysis, we propose the ProShrink algorithm and prove its convergence, and meanwhile we present detailed iteration schemes for solving models (1), (2), and (6). In section 4, we do sparse recovery experiments to demonstrate the benefit of adding box constrains.

## 2 Notation and important properties

In this paper, we restrict our attention onto two classes of intervals. The first class is:

$$T_1 = \{I : I = [c, \infty), c > 0, \text{ or } I = (-\infty, c], c < 0, \text{ or } I = [c, d], 0 < c < d \text{ or } I = [d, c], d < c < 0\};$$

The second class is

$$T_2 = \{I : I = [c, d], c < 0 < d\}.$$

The box constraint  $\mathcal{X}$  appeared in all models mentioned before is defined as  $\mathcal{X} = I_1 \times I_2 \times \cdots \times I_n$ , where  $I_i \in T_1 \cup T_2$ . Throughout this paper, we assume that  $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$ .

### 2.1 Basic concepts and properties

First, we introduce proximal point operator and its important properties.

**Definition 1.** Let  $f : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$  be a closed proper convex function. The proximal operator [6]  $\mathbf{prox}_f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined by

$$\mathbf{prox}_f(v) = \arg \min_x \left( f(x) + \frac{1}{2} \|x - v\|_2^2 \right). \quad (8)$$

Since the objective function is strongly convex and proper,  $\mathbf{prox}_f(v)$  is properly defined for every  $v \in \mathbb{R}^n$ . The following properties [6, 9] will be used in our analysis.

**Lemma 1.** Let  $f : \mathbb{R}^n \rightarrow R \cup \{+\infty\}$  be a closed proper convex function. Then, for all  $x, y \in \mathbb{R}^n$  the proximal operator  $\mathbf{prox}_{f(\cdot)}$  satisfies the followings:

1. *Firmly nonexpansive:*

$$\|\mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y)\|_2^2 \leq \langle x - y, \mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y) \rangle$$

2. *Lipschitz continuous:*  $\|\mathbf{prox}_{f(\cdot)}(x) - \mathbf{prox}_{f(\cdot)}(y)\|_2 \leq \|x - y\|_2$ .

If  $f$  is fully separable, meaning that  $f(x) = \sum_{i=1}^n f_i(x_i)$ , then

$$(\mathbf{prox}_{f(\cdot)}(x))_i = \mathbf{prox}_{f_i(\cdot)}(x_i). \quad (9)$$

Second, we need to introduce convex projected operator and projected subgradient to deal with box constraints.

**Definition 2** (convex projected operator).  $[x]_{\mathcal{X}}^+ := \arg \min_{y \in \mathcal{X}} \|x - y\|$ .

The following property of projected operator shall be often encountered in our deduction.

**Lemma 2.** For any interval  $I \in T_1 \cup T_2$ , we have that  $[\tau \cdot w]_I^+ = \tau \cdot [w]_{I/\tau}^+$  holds for arbitrary  $w \in \mathbb{R}$  and positive parameter  $\tau$ .

*Proof.* We begin with the definition of projected operator and derive that

$$[\tau \cdot w]_I^+ = \arg \min_{y \in I} \|y - \tau w\| = \arg \min_{y \in I} \|\tau^{-1} y - w\| = \tau \cdot \arg \min_{z \in I/\tau} \|z - w\| = \tau \cdot [w]_{I/\tau}^+.$$

This completes the proof.  $\square$

**Definition 3** (projected subgradient). Define  $\partial_{\mathcal{X}}^+ f(x) := \{g : g = x - [x - h]_{\mathcal{X}}^+, h \in \partial f(x)\}$  Without confusion, we also denote  $\partial_{\mathcal{X}}^+ f(x)$  by  $x - [x - \partial f(x)]_{\mathcal{X}}^+$ .

With projected subgradient, we can state a necessary and sufficient condition which guarantees a vector to be a minimizer to a class of constrained convex optimization problems.

**Lemma 3.** Let  $f(x)$  be proper convex and  $\mathcal{X}$  nonempty, closed, and convex. Then, we have

$$x^* \in \arg \min_{x \in \mathcal{X}} f(x) \Leftrightarrow 0 \in \partial_{\mathcal{X}}^+ f(x^*)$$

*Proof.* The following two facts will be used in our deduction:

**Fact 1.**  $z \in [x]_{\mathcal{X}}^+ \Leftrightarrow \langle z - x, z - y \rangle \leq 0, \forall y \in \mathcal{X}$ ;

**Fact 2.**  $\forall y \in \mathcal{X}, \exists h \in \partial f(x^*), \langle h, y - x^* \rangle \geq 0 \Leftrightarrow x^* \in \arg \min_{x \in \mathcal{X}} f(x)$ .

With these two facts, we derive that

$$0 \in \partial_{\mathcal{X},\tau}^+ f(x^*) \Leftrightarrow x^* \in [x^* - \partial f(x^*)]_{\mathcal{X}}^+ \quad (11a)$$

$$\Leftrightarrow \exists h \in \partial f(x^*), \text{ such that } x^* = [x^* - h]_{\mathcal{X}}^+ \quad (11b)$$

$$\Leftrightarrow \langle x^* - (x^* - h), x^* - y \rangle \geq 0, \forall y \in \mathcal{X} \quad (11c)$$

$$\Leftrightarrow \langle h, y - x^* \rangle \geq 0, \exists h \in \partial f(x^*), \forall y \in \mathcal{X} \quad (11d)$$

$$\Leftrightarrow x^* \in \arg \min_{x \in \mathcal{X}} f(x). \quad (11e)$$

This completes the proof.  $\square$

## 2.2 Projected shrinkage operator

In this part, we build an important formulation that links the proximal point operator and the projected shrinkage operation. In order to establish that formulation, we need two lemmas.

**Lemma 4.** *Let **shrink** be the shrinkage operator defined by  $\mathbf{shrink}(s) = \text{sign}(s) \max\{|s| - 1, 0\}$  and let  $I \in T_1$ . Then, we always have that*

$$[\mathbf{shrink}(q)]_I^+ = [q - \text{sign}(c)]_I^+ \quad (12)$$

holds for arbitrary  $q \in \mathbb{R}$ , where  $c$  appears in the definition of  $I$ .

*Proof.* Recall that  $I = [c, \infty)$  with  $c > 0$ , or  $I = (-\infty, c]$  with  $c < 0$ , or  $I = [c, d]$  with  $0 < c < d$ , or  $I = [d, c]$ , with  $d < c < 0$ . So it is easy to observe the following fact:

If  $c > 0$ , then  $[q + 1]_I^+ \equiv c$  for each  $q < -1$  and  $[0]_I^+ \equiv c$ ;

If  $c < 0$ , then  $[q - 1]_I^+ \equiv c$  for each  $q > 1$  and  $[0]_I^+ \equiv c$ .

Thus, together with the definition of the shrinkage operator, for  $c > 0$  we have that

$$[\mathbf{shrink}(q)]_I^+ = \begin{cases} [q - 1]_I^+, & q > 1 \\ [0]_I^+, & -1 \leq q \leq 1 \\ [q + 1]_I^+, & q < -1 \end{cases} \quad (13)$$

$$= \begin{cases} [q - 1]_I^+, & q > 1 \\ c, & q \leq 1 \end{cases} = [q - 1]_I^+ \quad (14)$$

and for  $c < 0$  have that

$$[\mathbf{shrink}(q)]_I^+ = \begin{cases} [q + 1]_I^+, & q < -1 \\ c, & q \geq -1 \end{cases} = [q + 1]_I^+. \quad (15)$$

On the other hand,  $[q - \text{sign}(c)]_I^+ = [q - 1]_I^+$  when  $c > 0$  and  $[q - \text{sign}(c)]_I^+ = [q + 1]_I^+$  when  $c < 0$ . So the relationship (12) holds.  $\square$

**Lemma 5.** *Let  $I \in T_1 \cup T_2$  and  $I_\tau(t) = \tau \cdot |t| + \delta_I(t)$  where  $\delta_I(\cdot)$  is the indicator function. Then, we always have that*

$$[\tau \cdot \mathbf{shrink}(\tau^{-1}q)]_I^+ = \mathbf{prox}_{I_\tau(\cdot)}(q) \quad (16)$$

holds for arbitrary  $q \in \mathbb{R}$ .

*Proof.* By the definition of proximal point operator, we derive that

$$t^* := \mathbf{prox}_{I_\tau(\cdot)}(q) = \arg \min_{t \in \mathbb{R}} I_\tau(t) + \frac{1}{2}(t - q)^2 = \arg \min_{t \in I} \tau \cdot |t| + \frac{1}{2}(t - q)^2. \quad (17a)$$

If  $I \in T_1$ , then  $|t| = \text{sign}(c) \cdot t$  and hence

$$t^* = \arg \min_{t \in I} \tau \cdot |t| + \frac{1}{2}(t - q)^2 = \arg \min_{t \in I} \tau \cdot \text{sign}(c) \cdot t + \frac{1}{2}(t - q)^2.$$

Applying Lemma 3 yields to  $t^* = [q - \tau \cdot \text{sign}(c)]_I^+$ . Together with Lemmas 2 and 4, we derive that

$$t^* = [\tau \cdot (\tau^{-1}q - \text{sign}(c))]_I^+ = \tau \cdot [\tau^{-1}q - \text{sign}(c)]_{I/\tau}^+ = \tau \cdot [\mathbf{shrink}(\tau^{-1}q)]_{I/\tau}^+ = [\tau \cdot \mathbf{shrink}(\tau^{-1}q)]_I^+.$$

So relationship (16) holds when  $I \in T_1$ .

If  $I \in T_2$ , then again invoking Lemma 3 yields to  $t^* \in [q - \tau \cdot \partial|t|_{t=t^*}]_I^+$ . Such  $t^*$  must be the unique solution to problem (17a) because its objective functions is strongly convex. Thus, it suffices to show that  $p(q) = [\tau \cdot \mathbf{shrink}(\tau^{-1}q)]_I^+$  satisfies the following inclusion:

$$t \in [q - \tau \cdot \partial|t|]_I^+.$$

Now, we check  $p(q)$  case-by-case:

**Case 1:**  $p(q) > 0$ . Since interval  $I = [c, d] \in T_2$  satisfies  $c < 0 < d$ , condition  $p(q) > 0$  implies  $\mathbf{shrink}(\tau^{-1}q) > 0$  and hence  $\tau^{-1}q > 1$ . Then, by Lemmas 2 and 4, and together with the definition of shrinkage operator we derive that

$$[q - \tau \cdot \partial\|p(q)\|_1]_I^+ = [q - \tau]_I^+ = [\tau \cdot (\tau^{-1}q - 1)]_I^+ = \tau \cdot [\tau^{-1}q - 1]_{I/\tau}^+ = \tau \cdot [\mathbf{shrink}(\tau^{-1}q)]_{I/\tau}^+ = [\tau \cdot \mathbf{shrink}(\tau^{-1}q)]_I^+ = p(q).$$

**Case 2:**  $p(q) < 0$ . Condition  $p(q) < 0$  implies  $\tau^{-1}q < -1$ . Then, similarly to the argument in Case 1, we have that

$$[q - \tau \cdot \partial\|p(q)\|_1]_I^+ = [q - \tau]_I^+ = [\tau \cdot (\tau^{-1}q + 1)]_I^+ = \tau \cdot [\tau^{-1}q + 1]_{I/\tau}^+ = \tau \cdot [\mathbf{shrink}(\tau^{-1}q)]_{I/\tau}^+ = [\tau \cdot \mathbf{shrink}(\tau^{-1}q)]_I^+ = p(q).$$

**Case 3:**  $p(q) = 0$ . Condition  $p(q) = 0$  implies  $-1 \leq \tau^{-1} \cdot q \leq 1$ . Then, noting the fact that  $[-1, 1] = \partial\|0\|_1$ , we have that

$$[q - \tau \cdot \partial\|p(q)\|_1]_I^+ = [q - \tau \cdot \partial\|0\|_1]_I^+ = \tau \cdot [\tau^{-1}q - \cdot \partial\|0\|_1]_{I/\tau}^+ \ni 0.$$

This completes the proof. □

Now, we are ready to build the most important formulation in this study.

**Corollary 1.** Define the projected shrinkage operator  $[\mathbf{shrink}(v)]_{\mathcal{X}}^+$  for a vector  $v \in \mathbb{R}^n$  via

$$([\mathbf{shrink}(v)]_{\mathcal{X}}^+)_i = [\mathbf{shrink}(v_i)]_{I_i}^+, i = 1, 2, \dots, n.$$

And let  $\mathcal{X}_\tau(x) = \tau \cdot \|x\|_1 + \delta_{\mathcal{X}}(x)$ . Then, it holds

$$\boxed{[\tau \cdot \mathbf{shrink}(\tau^{-1}v)]_{\mathcal{X}}^+ = \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(v)}. \quad (19)$$

*Proof.* Noting that  $\mathcal{X}_\tau(x) = \sum_{i=1}^n \tau|x_i| + \delta_{I_i}(x_i)$  and the property (9), together with Lemma 5, the conclusion follows. □

The significance of formulation (19) is two-fold: the expression based on proximal point operator will be used for convergence analysis; whilst that expressed by the projected shrinkage operator is for computational consideration due to its simplicity.

### 3 Projected shrinkage algorithm

In this section, we derive a Lagrange dual problem of (5) and the ProShrink algorithm for solving it. Following the line of proof thought in paper [17], we prove the convergence of both the primal and dual point sequences of the ProShrink algorithm.

#### 3.1 Lagrange dual analysis

The Lagrangian of the augmented convex model (5) is

$$L(x, y) = \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 + \langle y, b - Ax \rangle. \quad (20)$$

The Lagrange dual function is

$$D(y) = \underset{x \in \mathcal{X}}{\text{minimize}} L(x, y). \quad (21)$$

For any vector  $y \in \mathbb{R}^m$ , the  $x$ -minimization problem above is a strongly convex program and hence has a unique solution  $x^*(y)$  that satisfies

$$x^*(y) = \arg \min_{x \in \mathcal{X}} L(x, y) = \|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2 + \langle y, b - Ax \rangle \quad (22a)$$

$$= \arg \min_{x \in \mathcal{X}} \|x\|_1 + \frac{1}{2\tau} \|x - u - \tau \cdot A^T y\|_2^2 \quad (22b)$$

$$= \arg \min_{x \in \mathcal{X}} \tau \cdot \|x\|_1 + \frac{1}{2} \|x - u - \tau \cdot A^T y\|_2^2 = \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T y), \quad (22c)$$

where  $\mathcal{X}_\tau(x) = \tau \cdot \|x\|_1 + \delta_{\mathcal{X}}(x)$ . By formulation (19) in Corollary 1, we obtain

$$x^*(y) = \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T y) = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+. \quad (23)$$

Now,  $D(y) = L(x^*(y), y) = L([\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+, y)$ . Thus, we can write down the Lagrange dual problem of (5) as follows:

$$\underset{y}{\text{maximize}} L([\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+, y). \quad (24)$$

It is well known in convex analysis [10] that the dual objective function  $L([\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+, y)$  is gradient-Lipschitz-continuous due to the strong convexity of the primal objective function  $\|x\|_1 + \frac{1}{2\tau} \|x - u\|_2^2$ . And moreover, the gradient of dual objective function is given by

$$\nabla D(y) = b - Ax^*(y) = b - A[\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+$$

Each solution to the dual problem (24) can generate the unique solution to the primal problem (5) via formulation (23). This fact is stated in the following lemma.

**Lemma 6.** *Let  $x^*$  be the unique solution to problem (5) and  $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$ . Then the dual solution set to problem (24) is*

$$\mathcal{Y} = \{y : x^* = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+\}, \quad (25)$$

*which is nonempty and convex.*

*Proof.* By Lemma 1, for  $\forall y, \tilde{y} \in \mathbb{R}^m$  we have that

$$\begin{aligned} & \langle \nabla D(y) - \nabla D(\tilde{y}), y - \tilde{y} \rangle \\ &= \langle [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+ - [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T \tilde{y})]_{\mathcal{X}}^+, A^T y - A^T \tilde{y} \rangle \\ &= \tau^{-1} \langle \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T y) - \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T \tilde{y}), \tau \cdot A^T y - \tau \cdot A^T \tilde{y} \rangle \\ &\geq \tau^{-1} \|\mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T y) - \mathbf{prox}_{\mathcal{X}_\tau(\cdot)}(u + \tau \cdot A^T \tilde{y})\|_2^2 \geq 0, \end{aligned}$$

which implies that the dual objective function  $D(y)$  is convex. Thus, the dual solution set is

$$\mathcal{Y}' = \{y : \nabla D(y) = 0\} \quad (27a)$$

$$= \{y : A[\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y)]_{\mathcal{X}}^+ = b\}, \quad (27b)$$

which must be nonempty and convex by assumption  $\mathcal{X} \cap \{x : Ax = b\} \neq \emptyset$  and the convexity of  $D(y)$ . Now, it suffices to show  $\mathcal{Y} = \mathcal{Y}'$ . On one hand, we have  $\mathcal{Y} \subseteq \mathcal{Y}'$  since  $Ax^* = b$ . On the other hand, let  $y' \in \mathcal{Y}'$ , i.e.,  $y'$  is some dual solution. Then,  $x' = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y')]_{\mathcal{X}}^+$  is a primal solution and it must equal  $x^*$  by uniqueness. So  $y' \in \mathcal{Y}$  and hence  $\mathcal{Y}' \subseteq \mathcal{Y}$ , which completes the proof.  $\square$

### 3.2 Algorithm schemes

Applying the gradient iteration to the dual objective  $D(y)$  gives:

$$y^{k+1} = y^k + h(b - A[\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+), \quad (28)$$

where  $h > 0$  is the step size whose range shall be studied later for convergence. By setting  $x^{k+1} = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+$ , we obtain the equivalent iteration in the primal-dual form:

$$\begin{cases} x^{k+1} = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+ \\ y^{k+1} = y^k + h(b - Ax^{k+1}). \end{cases} \quad (29)$$

Because the projected shrinkage operator is involved, we call (29) projected shrinkage algorithm. Recall that the linearized Bregman (LBreg) algorithm [13, 2] has the following form:

$$\begin{cases} x^{k+1} = \tau \cdot \mathbf{shrink}(A^T y^k) \\ y^{k+1} = y^k + h(b - Ax^{k+1}). \end{cases} \quad (30)$$

Therefore, the ProShrink algorithm can be viewed as a generalization of the LBreg algorithm.

Applying Nesterov's accelerated scheme [7], we obtain an accelerated ProShrink algorithm with the following form:

$$\begin{cases} x^{k+1} = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+; \\ z^{k+1} = y^{(k)} + h(b - Ax^{k+1}); \\ \gamma_k = (\sqrt{\theta_k + 4} - \theta_k)/2; \\ \beta_{k+1} = (1 - \theta_k)\gamma_k, \theta_{k+1} = \theta_k \gamma_k; \\ y^{k+1} = z^{k+1} + \beta_{k+1}(z^{k+1} - z^k). \end{cases} \quad (31)$$

In addition, it is predictable that the adaptive restart technique developed in [8] can further accelerate the scheme (31); Such acceleration for the LBreg algorithm was observed in paper [16].

Now, let us return to models (1) and (2). Model (2) can be solved by ProShrink (29) or its acceleration (31) with  $u = 0$ . To solve model (1), we apply the proximal point algorithm and obtain a series of subproblems as follows:

$$z^{k+1} = \arg \min \{\|x\|_1 + \frac{1}{2\lambda_k} \|x - z^k\|^2, Ax = b, x \in \mathcal{X}\}$$

where  $\lambda_k$  are positive parameters. Each subproblem above can be well solved by ProShrink (29) as well. We write down the iteration scheme without detailed derivation:

$$\begin{cases} x^{i+1} = [\lambda_k \cdot \mathbf{shrink}(\frac{1}{\lambda_k} z^k + A^T y^i)]_{\mathcal{X}}^+ \\ y^{i+1} = y^i + h(b - Ax^{i+1}). \end{cases} \quad (32)$$

The subproblem can also be solved by the accelerated ProShrink scheme (31).

At last, the standard forward-backward splitting algorithm for model (6) is

$$x^{k+1} = \mathbf{prox}_{\mathcal{X}_{\gamma_k}(\cdot)} \left( x^k - \frac{\gamma_k}{\lambda} A^T (Ax^k - b) \right), \quad (33)$$

where  $\gamma_k$  are the step sizes. The main difficulty of the above iteration is to compute the proximal point operator of  $\mathcal{X}_{\gamma_k}(\cdot)$ . Utilizing formulation (19), this can be overcome and the iteration can be simplified into

$$x^{k+1} = [\gamma_k \cdot \mathbf{shrink}(\gamma_k^{-1} x^k - \lambda^{-1} A^T (Ax^k - b))]_{\mathcal{X}}^+. \quad (34)$$

### 3.3 Convergence analysis

In this part, we prove the convergence of primal sequence  $\{x^k\}$  and dual sequence  $\{y^k\}$  in iteration (29).

**Theorem 1.** *Set step size  $h \in (0, \frac{2}{\tau\|A\|^2})$  and  $y^0 = 0$  in iteration (29). Let  $x^*$  be the unique minimizer to problem (2) and  $\mathcal{Y}$  be the solution set to problem (24). Then,  $\lim_{k \rightarrow +\infty} x^k = x^*$ , and there exists a point  $\bar{y} \in \mathcal{Y}$  such that  $\lim_{k \rightarrow +\infty} y^k = \bar{y}$ .*

This theorem can be proved in the same manner as that in paper [17]. For completeness, we provide a proof below.

*Proof.* Let  $\hat{y} \in \mathcal{Y}$ . By Lemma 6, we have  $x^* = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T \hat{y})]_{\mathcal{X}}^+$ . Together with  $x^{k+1} = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+$  and Lemma 1 and Corollary 1, we derive

$$\langle A^T y^k - A^T \hat{y}, x^{k+1} - x^* \rangle \quad (35a)$$

$$= \langle A^T y^k - A^T \hat{y}, [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^k)]_{\mathcal{X}}^+ - [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T \hat{y})]_{\mathcal{X}}^+ \rangle \quad (35b)$$

$$= \langle A^T y^k - A^T \hat{y}, \mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T y^k) - \mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T \hat{y}) \rangle \quad (35c)$$

$$\geq \tau^{-1} \cdot \|\mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T y^k) - \mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T \hat{y})\|_2^2 \quad (35d)$$

$$= \tau^{-1} \cdot \|x^{k+1} - x^*\|_2^2 \quad (35e)$$

Using this inequality, we have

$$\|y^{k+1} - \hat{y}\|_2^2 = \|y^k - \hat{y} + h(b - Ax^{k+1})\|_2^2 \quad (36a)$$

$$= \|y^k - \hat{y} + h(Ax^* - Ax^{k+1})\|_2^2 \quad (36b)$$

$$= \|y^k - \hat{y}\|_2^2 - 2h \langle A^T y^k - A^T \hat{y}, x^{k+1} - x^* \rangle + h^2 \|Ax^* - Ax^{k+1}\|_2^2 \quad (36c)$$

$$\leq \|y^k - \hat{y}\|_2^2 - 2h\tau^{-1} \|x^{k+1} - x^*\|_2^2 + h^2 \|A\|^2 \|x^{k+1} - x^*\|_2^2 \quad (36d)$$

$$= \|y^k - \hat{y}\|_2^2 - h(2\tau^{-1} - h\|A\|^2) \|x^{k+1} - x^*\|_2^2. \quad (36e)$$

Therefore, under the assumption  $0 < h < \frac{2}{\tau\|A\|^2}$  we can make the following claims:

**claim 1:**  $\|y^{k+1} - \hat{y}\|_2$  is monotonically nonincreasing in  $k$  and thus converges to a limit;

**claim 2:**  $\|x^{k+1} - x^*\|_2$  converges to 0 as  $k$  tends to  $+\infty$ , i.e.,  $\lim_{k \rightarrow +\infty} x^{k+1} = x^*$ .

From claim 1, it follows that  $\{y^k\}$  is bounded and thus has a converging subsequence  $y^{k_i}$ . Let  $\bar{y} = \lim_{i \rightarrow \infty} y^{k_i}$ . By the Lipschitz continuity of proximal point operator in Lemma 1 and Corollary 1, we have

$$\begin{aligned} x^* &= \lim_{i \rightarrow \infty} x^{k_i+1} = \lim_{i \rightarrow \infty} [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T y^{k_i+1})]_{\mathcal{X}}^+ \\ &= \lim_{i \rightarrow \infty} \mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T y^{k_i+1}) = \mathbf{prox}_{\mathcal{X}_{\tau}(\cdot)}(u + \tau \cdot A^T \bar{y}) = [\tau \cdot \mathbf{shrink}(\tau^{-1}u + A^T \bar{y})]_{\mathcal{X}}^+, \end{aligned}$$

so  $\bar{y} \in \mathcal{Y}$  by Lemma 6. Recall  $\hat{y} \in \mathcal{Y}$  is arbitrary. Hence, claim 1 holds for  $\hat{y} = \bar{y}$ . If  $\{y^k\}$  had another limit point, then  $\|y^{k+1} - \bar{y}\|_2$  would fail to be monotonic. So,  $y^k$  converges to  $\bar{y} \in \mathcal{Y}$  (in norm).  $\square$



## 4 Numerical experiment

In the section, we do sparse recovery experiments to demonstrate that adding box constraints can help improve recovery of sparse signals considerably. It was shown in [5] when the augmented parameter  $\tau \geq 10\|x\|_\infty$ , the augmented  $\ell_1$ -norm model (4) is equivalent to classical basis pursuit (3) if the sensing matrix  $A$  satisfies certain properties such as null-space property, or restricted isometry property. So we only test models (4) and (2) to observe possible advantages of adding box constraints. In the test, model (4) was solved by the LBreg algorithm and model (2) by the ProShrink algorithm.

We used 100 random pairs  $(A, x)$  with matrices  $A$  of size  $200 \times 400$  and vectors  $x$  with 400 entries, out of which  $s$  were nonzero entries set to  $\pm 1$  uniformly randomly for  $s = 1, 2, 3, \dots, 80$ . Each entry of the sensing matrix  $A$  was sampled independently from the standard Gaussian distribution. Thus,  $b = Ax$  are given vectors. A relative error of  $10^{-12}$  was considered as an exact recovery; the relative error is defined as  $\frac{x - x_o}{x}$  where  $x_o$  is finally generated by the LBreg or the ProShrink algorithms. The box-constrained set  $\mathcal{X}$  for the ProShrink algorithm was set as  $[-1, 1]^{400}$ .

We plot the exact recovery rate via sparsity levels in Figure 1 from which we see that ProShrink performs remarkably better than LBreg as the sparse level increases. More precisely, when the sparse level is low, both LBreg and ProShrink can well recover sparse signals; but when the sparse level becomes high, the recovery rate by LBreg is worse than that by ProShrink that indicates adding box constraints to the augmented  $\ell_1$ -norm model (4) indeed improves the recovery rate.

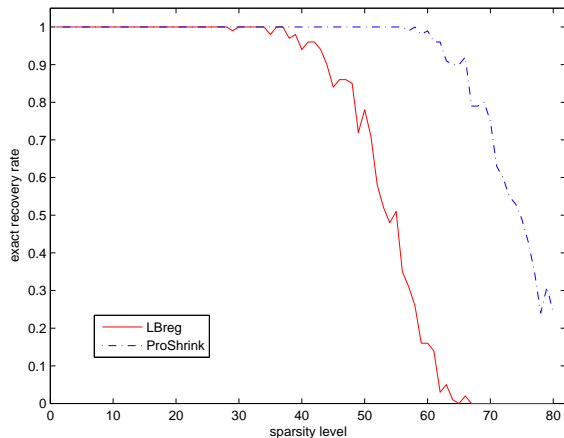


Figure 1: Comparison of augmented  $\ell_1$  norm models with or without box constraints for sparse recovery (correspond to the ProShrink and the LBreg algorithms separately).

## 5 Conclusion

In this paper, we proposed the projected shrinkage algorithm for boxed-constrained  $\ell_1$ -minimization. The most important factor in our study should be the deduction of formulation (19) that establishes the relationship between projected shrinkage operator and proximal point operator. Numerically, we demonstrated that adding box constraints to classical  $\ell_1$ -minimization can obtain better performance. However, giving theoretical explanation for this phenomenon is open. We leave it for future work.

## Acknowledgements

We would like to thank Professor Wotao Yin (UCLA) for his comments and suggestion on numerical verification and Professor Jian-Feng Cai (Iowa U) for his insight of the projected shrinkage operator.

## References

- [1] A. M. Bruckstein, M. Elad and M. Zibulevsky, *On the Uniqueness of Nonnegative Sparse Solutions to Underdetermined Systems of Equations*, IEEE Trans. on Information Theory, 54(11):4813-4820, 2008.
- [2] J.-F. Cai, S. Osher, and Z. Shen, *Linearized Bregman iterations for compressed sensing*, Mathematics of Computation, 78(267):1515-1536, 2009.
- [3] S. Chen, D. Donoho, and M. Saunders, *Atomic decomposition by basis pursuit*, SIAM Review, 43(1):129-159, 2001.
- [4] D. L. Donoho and J. Tanner, *Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming*, Proceedings of the National Academy of Sciences, 102(27):9446-9451, 2005.
- [5] M. J. Lai and W. Yin, *Augmented  $\ell_1$  and nuclear-norm models with a globally linearly convergent algorithm*, SIAM. J. Imaging Sci., 6(2):1059-1091, 2013.
- [6] J. J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris Ser. A Math. 255:2897-2899, 1962.
- [7] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Kluwer Academic Publishers, 2004.
- [8] B. O'Donoghue and E. Candès, *Adaptive restart for accelerated gradient schemes*, To appear in Foundations of Computational Mathematics, 2012.
- [9] N. Parikh, and S. Boyd, *Proximal algorithm*, Foundations and Trends in optimization, 1-96, 2013.
- [10] R. T. Rockafellar, *Convex analysis*, Princeton, 1970.
- [11] M. Takáč, J. Mareček, and P. Richtárik, *Inequality-Constrained Matrix Completion: Adding the Obvious Helps!* arXiv:1408.2467v1 [math.OC] 11 Aug 2014.
- [12] N. Vaswani and W. Lu, *Modified-CS: Modifying Compressive Sensing for Problems with Partially Known Support*. 2008.
- [13] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing*, SIAM Journal on Imaging Sciences, 1(1):143-168, 2008.
- [14] H. Zhang, L. Cheng, and W. Zhu, *A lower bound guaranteeing exact matrix completion via singular value thresholding algorithm*, Applied and Computational Harmonic Analysis, 31(3):454-459, 2011.
- [15] H. Zhang, J.-F. Cai, L. Cheng, and J. Zhu, *Strongly convex programming for exact matrix completion and robust principal component analysis*, Inverse Problem and Imaging, 6(2): 357-372, 2012.
- [16] H. Zhang and W. Yin, *Gradient methods for convex minimization: better rates under weaker conditions*, UCLA CAM Report(13-17), 2013.
- [17] H. Zhang, L. Z. Cheng, and W. Yin, *Dual algorithm for a class of augmented convex signal recovery models*, Communications in Mathematical Sciences, 13(1): 103-112, 2015.